Trade-Off Sensitive Experimental Design: A Multicriterion, Decision Theoretic, Bayes Linear Approach

Malcolm Farrow
Newcastle University, UK

Michael Goldstein
University of Durham, UK

Abstract

We show how mutually utility independent hierarchies, which weigh the various costs of an experiment against benefits expressed through a mixed Bayes linear utility representing the potential gains in knowledge from the experiment, provide a flexible and intuitive methodology for experimental design which remains tractable even for complex multivariate problems. A key feature of the approach is that we allow imprecision in the trade-offs between the various costs and benefits. We identify the Pareto optimal designs under the imprecise specification and suggest a criterion for selecting between such designs. The approach is illustrated with respect to an experiment related to the oral glucose tolerance test.

Keywords: Imprecise utility; Multi-attribute utility; Pareto optimality; Oral glucose tolerance test.

1 Introduction

This paper is concerned with the decision theoretic approach to the design of experiments. In principle, the choice of an experimental design is a decision problem. We have a number of possible choices of design (decisions). Each choice has certain benefits, typically derived from expected gains in knowledge, and incurs expenditure, typically relating to both resource and ethical costs. Choosing a good design is a hard decision problem, which would benefit, in two fundamental ways, from appropriate methodology, in which all costs and benefits were quantitatively assessed, combined and jointly optimised. Firstly, the **formulation** of design choice as a decision problem, by making all relevant judgments explicit, would bring a clarity to the process. Such a quantitative formulation would make explicit the implied trade-offs between costs and benefits which underlie the choice of design for each of the participants, namely the experimenters, the subjects in the experiment, the ethics committee, the funding body and so forth. Secondly, the **solution** of the resulting decision problem would identify the best choice of design, or, more in the spirit of decision analysis, the collection of designs which are almost optimal would identify the pragmatic range of design choices that merit serious consideration.

Among the problems that we must address in developing a decision theoretic approach to design are the following: (i) **costs:** resource usage, ethical costs, and so forth may be hard to quantify; (ii) **benefits:** potential gains in knowledge from the experiment may be even harder to quantify; (iii) **trade-offs:** experimenters may find it very difficult to weigh the relative importance of the different costs, and even harder to weigh such costs against the potential benefits of the experiment; (iv) **prior knowledge**: this is hard to quantify for complex problems; (v) **design**: even were we able to specify fully all of the above ingredients, for example within a full Bayes framework, choosing the optimal design would be an extremely difficult computing problem.

While there is plenty of theoretical work on Bayes designs, (see, for example, the review in Chaloner and Verdinelli, 1995), there is little Bayes (or any other!) methodology which, in practice, offers guidance on decision theoretic design for complex multivariate experiments with realistic cost structures. Much of the Bayesian design literature is concerned with one or other of two aspects of the problem. The first, which we might term the *design-point selection problem*, is concerned with finding an "optimal" design, often specified by a design matrix, given a fixed total sample size. Most literature on the second aspect, the *sample size* problem, involves simple designs and, usually, a simple trade-off between the cost of experimental units and information gain.

In the design-point selection literature, by relaxing the requirement for the numbers of observations at different design points to be integers, the problem, at least in linear models, can be made one of selecting the *proportions* of the total sample to be allocated to each design point. Thus the question of choosing a sample size is avoided. Moreover, usually no allowance is made for differences in costs between observations at different design points. The optimality criterion is often the Bayesian version of a criterion such as A- or D-optimality. These two and some others can be justified as maximising the expectation of a utility function which values information gain in some sense. For example, Sebastiani and Wynn (2000) use maximum entropy sampling which, in the normal linear model, leads to Bayesian D-optimality. Recent work in this area has included the combination of several information-gain criteria, for example to measure learning about different possible models. Examples of this include Clyde and Chaloner (1996) and Dette and Franke (2000). The criteria are usually combined as a weighted sum or one is optimized subject to constraints on the values attained for the others. In this paper we combine an information-gain utility with utilities for costs. The information-gain utility involves combining, not necessarily additively, a number of sub-utilities, each of which is related to the utility leading to the Bayesian A-optimality criterion. Verdinelli and Kadane (1992) use a utility which combines information gain, in the D-optimality form, with the total value of the dependent variable in an experiment.

The Bayesian sample-size literature generally involves maximising the expectation of a utility which involves the cost of making observations and either the consequence of some action following the experiment or a measure of information gain. Lindley (1997), Bernardo (1997) and Adcock (1997) give a useful introduction to the area. Recent examples include Walker (2003), whose utility involves the consequence of an action such as choice of treatment, for the case of a simple random

sample, and Gittins and Pezeshk (2000), who propose both a "public health" utility, involving both the benefit of a new treatment and the number who will use it as well as costs, and a "commercial" utility, for the case of a clinical trial to compare two treatments. For a Bayes linear approach to finding sample sizes to obtain specified reductions in variance, see Goldstein and Wooff (1997) and Shaw and Goldstein (1999).

Tan and Smith (1998) consider the use of realistic utility functions for both sample-size determination and allocation to treatments in a two-treatment clinical trial, with a cost-benefit trade-off in which benefit is related to "success" of the trial. Claxton *et al.* (2000) discuss choice of treatment (not design) with "cost-benefit" trade-off.

This paper has two parts. In the first part, we develop a multi-criterion decision theoretic approach to the design of experiments. We use a flexible cost utility structure which allows costs of several types and differences in costs between design points. We allow multivariate observations and use a correspondingly flexible information-gain utility. To provide a focus for our discussion, we discuss a multivariate, grouped, repeated measurement study in a medical context, with two types of cost (financial, ethical), which must be weighed against the potential gains from the experiment.

While decision theory addresses the choices of the individual, experimental design typically needs to reconcile the preferences of many protagonists. Further, even in fairly simple examples, a single individual may not be able to specify precisely the relative importance of the various costs and benefits. Therefore, one of the most difficult steps in many decision analyses is the quantification of the relative importance of the different attributes of the design, due either to the unwillingness or inability on the part of the experimenter to specify certain risk trade-offs or from disagreement on priorities within a committee with responsibility for the design. In the second part of the paper, we therefore address this issue by developing an approach which explicitly allows for and incorporates such imprecision in the trade-offs between the different costs and benefits of the experiment.

In Section 2 of this paper we introduce our illustrative example. In Section 3 we introduce a suitable structure for utility functions in experimental design problems and, in particular, a method of assigning utilities to the gains in knowledge obtained by the experiments. In Section 4 we describe how an optimal design may be selected. In Section 5 we generalise the method by allowing imprecision in the trade-offs between the attributes of the utility. This leads to the selection of a set of Pareto optimal designs as shown in Section 6. A criterion for selecting between the Pareto optimal designs is suggested in Section 7. Finally, in Section 8, we extend the method to allow a more general form of utility function.

2 Example: The oral glucose tolerance test

We shall consider an experiment, actually conducted some years ago, concerning the oral glucose tolerance test (O.G.T.T.) which is used to help measure the severity of disease suffered by a diabetic patient. After a period without food, patients are given glucose orally. Blood samples are taken immediately before the glucose is given and at intervals thereafter. The concentration of glucose in each sample is determined and, roughly speaking, the quicker the glucose concentration returns to its fasting level the more healthy is the patient. High fasting levels of glucose, or slow return to fasting levels after taking glucose, are suggestive of diabetes. For example, the World Heath Organisation diagnostic levels were that a fasting glucose level above 7 mmol/litre or a two hour level above 10 mmol/litre suggests diabetes, while a fasting level below 7 mmol/litre and a two hour level between 7 and 10 mmol/litre suggest "impaired glucose tolerance" (World Health Organisation, 1980). One of the main reasons for the study was that the experimenters believed that methods of interpretation of O.G.T.T. results designed for young patients might need to be altered to suit the elderly because of a generally slower response rate in old people. The groups of elderly subjects in the study included Group 1: "Well-controlled" (diabetes controlled by diet with or without drugs but not insulin), and Group 2: "Insulin dependent" (the most severe form, requiring insulin treatment). (To simplify our account, we accept the clinicians' classifications at face value.)

Blood samples were taken immediately before the glucose was given and at intervals thereafter.

From each sample, the concentrations of glucose and, in addition, C-peptide could be determined. (See also Wickramasinghe et al.,1992, Farrow and Leyland, 1991, and Farrow and Goldstein, 1992). The experiment had two aims. Firstly, in order to improve the interpretation of test results for elderly patients and perhaps shorten test duration for such patients, information was required on typical responses to the O.G.T.T. from elderly patients with different degrees of diabetes. Secondly the investigators wished to see whether C-peptide measurements, in addition to blood glucose, provided extra diagnostic information, as C-peptide concentration in the blood is thought to indicate insulin production.

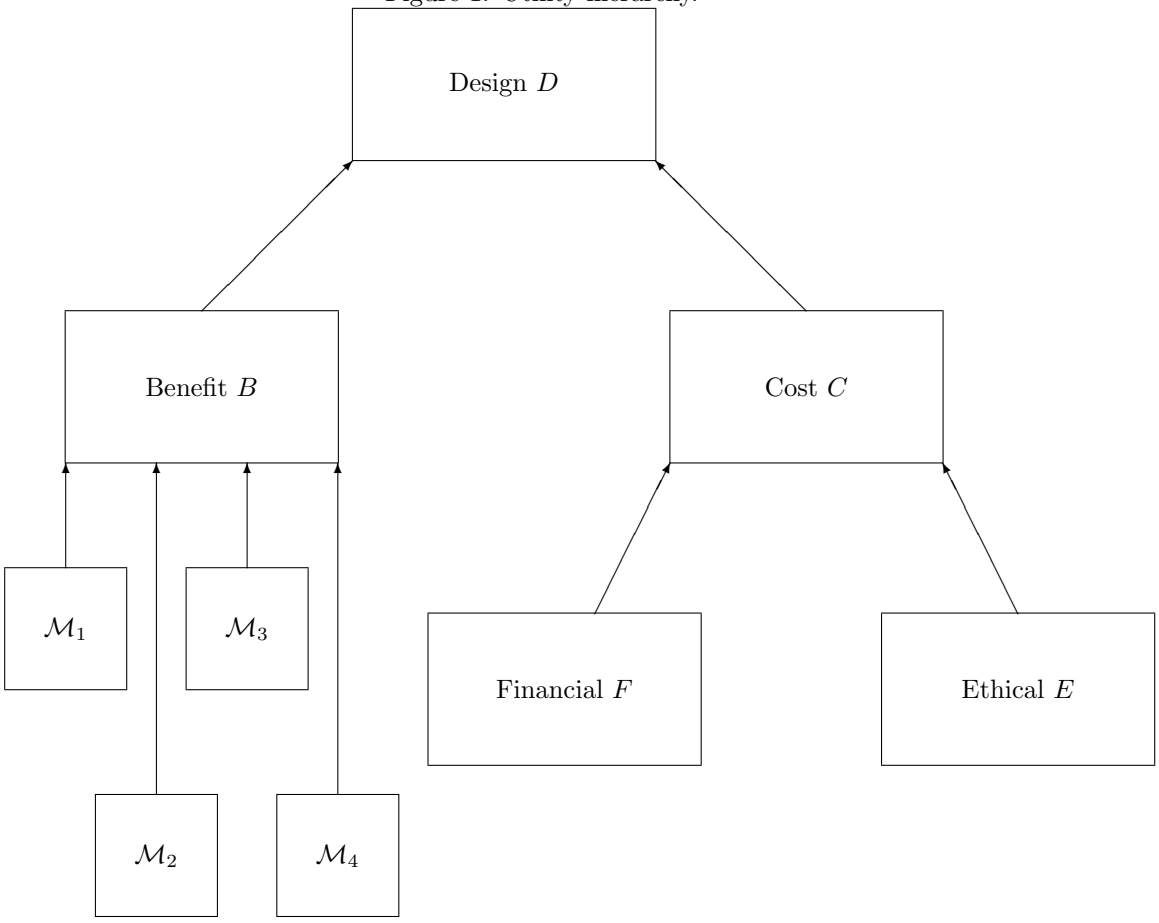
There are two types of cost in this experiment. Firstly, there are financial costs, both in taking each blood sample and also the laboratory costs for each determination of glucose or C-peptide concentration. It costs much more to determine a C-peptide concentration than a glucose concentration. Further, there are staff costs involved in the care and administration for each patient and these increase with the duration of the test and the severity of illness. Secondly, there are ethical costs, which arise not only from the inconvenience and discomfort, but also because there may be damage to a diabetic patient's health, from fasting, taking glucose and fasting again. These costs greatly increase with the severity of the illness and the duration of the test. They must be weighed against the benefits of the experiment which derive from our gain in knowledge.

To simplify our account, we restrict our attention to Groups 1 and 2 and to the baseline and 60 minute values and we suppose that we always take the cheaper observation, namely the glucose level, for each patient, making the more expensive, C-peptide, observation on a subset of the patients. Thus our design choices are to determine four numbers n_{gv}, where n_{gv} is the number of patients chosen from group g, where $g = 1, 2$, for which variables v are observed where $v = 1$ labels the subset on which only glucose measures are taken, and $v = 2$ labels the subset on which both glucose and C-peptide measures are taken. We will denote a particular design choice by the quadruple $(n_{12}, n_{22}, n_{11}, n_{21})$.

If, as here, financial and ethical costs are of serious concern, then decision analysis is the only approach which allows us to weigh the benefits of different experiments against the costs in a manner which allows us to identify the design of maximal value. In order to carry out such an analysis, we must (i) formulate a utility function on experimental costs and outcomes, (ii) quantify all uncertainties and (iii) optimise expected utility over all choices of design. Of course, if we do not use decision analysis, we will still have to choose a design, but without a formal approach for weighing costs and benefits the choice will be much more difficult and, to a large extent, arbitrary. In the following sections, we shall show how to set up and analyse experiments such as these within the decision formalism, and then introduce certain extensions to the standard methodology which are useful for dealing with problems involving risk trade-offs that are considered difficult to specify.

The first step is to construct the utility function. To motivate our general discussion, observe that it is often valuable to think of the utility in the form of a hierarchy. See figure 1. In the present example, the top level node is the overall utility of the designed experiment. We may view this utility as having two attributes, namely Costs and Benefits. So on the second level we have a Cost node and a Benefit node, both leading into the Overall node. The Cost node itself has two attributes, namely Ethical and Financial costs, and so, on the third level, each has a node, both leading into the Cost node. The Benefit node similarly has various attributes, all arising from gains in information about the various patient responses in the two groups. There are many ways that we may choose to represent such benefits, depending on the context in which we plan to use our results. In our account, we will suggest a simple choice for benefits, which is sufficient to illustrate the general approach, namely to suppose that there are four attributes for the benefit comprising improved information about each of the following collections of quantities: $\mathcal{M}_1$ fasting levels for C-peptide; $\mathcal{M}_2$ fasting levels for glucose; $\mathcal{M}_3$ changes in C-peptide levels over one hour; $\mathcal{M}_4$ changes in glucose levels over one hour. For example, $\mathcal{M}_1$ represents gains in information about the joint distribution of fasting C-peptide levels both for individuals who are "Well controlled" and individuals who are "Insulin dependent". We will clarify the precise way in which this benefit is quantified in later sections. Our utility hierarchy is therefore completed by adding nodes representing each of $\mathcal{M}_1$ to $\mathcal{M}_4$ at the third level, entering the Benefit node.

Figure 1: Utility hierarchy.



3 Utility hierarchies for experimental design

3.1 Mutually utility independent hierarchies

We now explain the formal properties that we require of a utility hierarchy, in order that it should be amenable to formal specification and analysis, and construct appropriate utility functions for experimental design. In this section we consider a general class of multi-attribute utility functions. Attributes $\boldsymbol{Y} = (Y_1, ..., Y_k)$ are *utility independent* of the attributes $\boldsymbol{Z} = (Z_1, ..., Z_r)$ if conditional preferences over lotteries with differing values of $\boldsymbol{Y}$ but fixed values, $\boldsymbol{z}$, of $\boldsymbol{Z}$, do not depend on the particular choice of $\boldsymbol{z}$. Attributes $\boldsymbol{X} = (X_1, ..., X_s)$ are *mutually utility independent* if every subset of $\boldsymbol{X}$ is utility independent of its complement. If attributes $\boldsymbol{X}$ are mutually utility independent, then the utility function for $\boldsymbol{X}$ must be given by the *multiplicative form*, $(1 + kU(\boldsymbol{X})) = \prod_{i=1}^{s}(1 + ka_iU_i(X_i))$, or the *additive form*, $U(\boldsymbol{X}) = \sum_{i=1}^{s} a_iU_i(X_i)$, (see Keeney and Raiffa, 1976), where $U_i(X_i)$ is a conditional utility function for attribute X_i, namely an evaluation of the utility of X_i for fixed values of the other attributes. The coefficients in these equations are the *trade-off parameters*; the a_i reflect the relative importance of the attributes and k reflects the degree to which rewards may be regarded as complementary, if $k > 0$, or as substitutes, if $k < 0$ (Keeney and Raiffa, 1976, section 5.4.5). Keeney and Raiffa (1976) also describe the idea of a hierarchy of utilities, as follows. We form an overall multi-attribute utility from marginal utilities for the various attributes by a hierarchical structure which may be represented as a graph in which each attribute of the problem is represented as a node. Some of the attributes are themselves comprised of sub-attributes and, in the graphical representation, each attribute node is joined to each of the constituent sub-attributes by an arc. An example of such a structure is given as figure 1. The overall utility is constructed as follows. For each attribute with no sub-attributes, we specify a utility function. At each other node in the structure, we construct the utility by merging the utilities of all of the constituent sub-attributes. Eventually, one overall utility function is formed corresponding to the node at the head of the graph. If, at each node, we have mutual utility independence for the utilities combined at that node, then we term such a utility function a *Mutually Utility Independent Hierarchic (MUIH)* utility. Thus, in a MUIH utility, at each node we combine utilities using either the multiplicative or additive form.

In our utility hierarchy we consider the overall utility node to be at the "top" level and the predecessors of a node to be at "lower" levels. We refer to the nodes corresponding to the individual attributes, that is nodes which have no predecessors, as *marginal nodes*. We refer to a direct predecessor of a node as a *parent* and a direct successor as a *child*. For each node i, we denote by $N(i)$, the *sub-hierarchy* under i, where $N(i)$ is the set of nodes containing i and all of its predecessors. We divide the child nodes in the hierarchy into three types: *additive*, *binary* and *multiplicative*. In an *additive node* utilities are combined using the additive form giving

$$U = \sum_{i=1}^{s} a_iU_i \tag{1}$$

with $\sum_{i=1}^{s} a_i \equiv 1$ and $a_i > 0$ for $i = 1, \ldots, s$. In a *binary node*, with parameters a_1, a_2, h, precisely two utilities are combined and we rescale the combined utility as

$$U = a_1U_1 + a_2U_2 + hU_1U_2 \tag{2}$$

where $0 < a_i < 1$, $-a_i \leq h \leq 1 - a_i$ and $a_1 + a_2 + h = 1$. In a *multiplicative node* more than two utilities are combined and the parameter k in the multiplicative form may be nonzero. We rescale the utility so that

$$U = \frac{\prod_{i=1}^{s}(1 + ka_iU_i) - 1}{\prod_{i=1}^{s}(1 + ka_i) - 1} \tag{3}$$

where $a_1 \equiv 1, k > -1$ and, for $i = 1, \ldots, s$, we have $a_i > 0$ and $ka_i > -1$.

The above categorisation embraces all of the different specifications that we may make which are consistent with the property of utility independence. (Binary nodes are the special case of

multiplicative nodes where $s = 2$. Equation (2) is derived by setting $s = 2$ and $h = ka_1a_2$ in (3). We distinguish between binary and multiplicative nodes only because some results may be stated more simply in the binary case.) For some problems, it will be sufficient to work only with additive nodes. However, we gain much flexibility by allowing the more general forms, as this allows us to consider problems where, while the attributes are utility independent, they still complement or substitute for each other. See the discussion in Keeney and Raiffa (1976).

For each child node i, we denote by $\boldsymbol{\phi}_i = (\phi_{i1}, \ldots, \phi_{im(i)})$ the collection of trade-off parameters which determine how the parent utilities at node i are combined to give the value at the child node. Thus, each ϕ_{ij} corresponds to an a_i in (1) an a_i or h term in (2), or an a_i or k in (3). If there are n child nodes, then we denote by $\boldsymbol{\theta} = (\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_n)$ the collection of all the trade-off parameters in the hierarchy. As the marginal utility at each marginal node is expressed in a utility scale, we norm all the marginal utilities to lie between 0, the worst outcome that we shall consider for the problem, and 1, the best outcome. The effect of the scalings that we have chosen for additive, binary and multiplicative nodes is that, at each node i in the hierarchy, the utility is 1 for the outcome C_i when all marginal predecessor nodes have utility 1, and is zero for the outcome c_i when all marginal predecessor nodes have utility zero. Therefore, a utility value of u at node i may always be interpreted as the utility of a gamble giving C_i with probability u and c_i with probability $1-u$, irrespective of the chain of trade-off parameters in the hierarchy. This utility scale is termed the **standard scale** for the hierarchy. Throughout this paper, all utilities are assumed to be on the standard scale.

3.2 Costs and benefits

We now construct MUIH hierarchies for experimental design. Our overall utility U(D) for choosing a design D depends on the costs, C, and the benefits, B. If costs and benefits can be taken to be mutually utility independent, then our overall utility, U_d, for a design is a binary node (2) with parents U_c, U_b, the marginal utilities for cost and benefit, scaled to [0,1], with parameters a_c, a_b, h.

$$U_d = a_c U_c + a_b U_b + h_d U_c U_b \tag{4}$$

We now separate U_c and U_b into contributory attributes. As an illustration, suppose that U_c depends both on ethical costs, E, and financial costs, F, which we judge to be utility independent. Then

$$U_c = a_e U_e + a_f U_f + h_c U_e U_f \tag{5}$$

where U_e and U_f are utilities for ethical and financial costs respectively, placed on a $[0, 1]$ scale by setting each cost utility to one for the null experiment, i.e. no experiment at all, and to zero for the largest cost which might be tolerated.

In some circumstances, it will be reasonable to construct benefit utilities which relate directly to the effectiveness of potential future treatments in a well-defined clinical context, e.g. Gittins and Pezeshk (2000), Walker (2003). Such utility specifications will typically be highly problem specific and usually will require great care in taming the computational complexity required to identify good designs, as Bayesian design choice is a notoriously computer intensive problem, often requiring simulations (e.g. Müller, 1999). In many cases, however, treatment decisions will not follow directly from the individual study, which, as in our example, is part of an ongoing series of investigations which will in combination identify treatment regimes. In such cases, the main benefits from a medical experiment derive from our gain in knowledge. This might lead to the development of better diagnostic or therapeutic methods, it might show that the test or treatment was useless or it might lead to further experiments which are eventually fruitful in some currently unsuspected way. Therefore, a natural way to express the benefit is to quantify the change in information, relative to our current state of knowledge. This view, of course, is also the justification for the use of criteria such as Bayesian A- and D-optimality which are widely used in the Bayesian design literature (Chaloner and Verdinelli, 1995). We may then weigh the value of possible gains in knowledge against the various costs. (Additional benefits, for example direct benefit to patients in

a trial from the new treatment, may be handled analogously to the costs.) We now develop such a benefit utility in a flexible form which leads to a tractable design methodology.

3.3 Mixed Bayes linear utilities for gains in knowledge

Design calculations for full Bayes analyses are notoriously expensive, as, for each choice of design, the reduction in uncertainty must be assessed for each possible sample that could be obtained. Therefore, we instead choose to assess the reduction in uncertainty from a Bayes linear viewpoint. This approach may be viewed either as applying when the underlying uncertainties are roughly Gaussian, or as providing a general, simple and tractable upper bound approximation to the preposterior expectation of the posterior variance given certain simple moment assessments, or as the appropriate form of analysis when we are only prepared to make a partial prior belief specification in the form of first and second moments. (See Goldstein, 1999, and Farrow and Goldstein, 1993). However, provided the experiment is sufficiently simple for a full Bayes preposterior design calculation to be tractable, then the methodology that we develop below for design choice will apply equally for any such Bayes analysis. We now describe the Bayes linear utility, which we build in three stages.

(i) Consider an experiment in which we learn about a single quantity, X. Suppose that, from design D, we observe data d, and calculate the Bayes linear adjusted expectation, $\mathrm{E}_d(X)$ for X. As the Bayes linear approach takes expectation as primitive it is natural and tractable to consider a basic attribute of the experiment to be the magnitude of the difference between $\mathrm{E}_d(X)$ and the actual value of X, namely $d(X) = X - \mathrm{E}_d(X)$, assessed as $(X - \mathrm{E}_d(X))^2$ (see, for example, de Finetti (1974) in which the preferred choice of penalty of this type is taken as the operational definition for expectation). A simple benefit utility, constructed from this attribute, is given by the expectation of this quantity, namely $\mathrm{E}(d^2(X)) = \mathrm{var}_d(X)$, the adjusted variance of X given d. We scale so that the utility of carrying out a **precise experiment**, i.e. one which will reveal exactly the value of X, is 1, and the utility of carrying out no experiment is 0, so that $U(X) = 1 - \mathrm{var}_d(X)/\mathrm{var}(X)$.

(ii) Now suppose that we learn about a collection $\boldsymbol{X} = (X_1, ..., X_m)$, where all linear combinations of $\boldsymbol{X}$ are considered equally important. Then, a corresponding collection of attributes which express our gain in knowledge is the vector $d(\boldsymbol{X}) = \boldsymbol{X} - \mathrm{E}_d(\boldsymbol{X})$, where $\mathrm{E}_d(\boldsymbol{X})$ is the vector of adjusted expectations for the elements of $\boldsymbol{X}$ after observing data d. If $(X_1, ..., X_m)$ are uncorrelated, then a corresponding representation of the joint utility is the additive form, namely $U(\boldsymbol{X}) = m^{-1} \sum_{i=1}^{m} U(X_i)$, where each $U(X_i) = 1 - \mathrm{var}_d(X_i)/\mathrm{var}(X_i)$. If $(X_1, ..., X_m)$ are correlated, then we may apply this to the principal components of $\boldsymbol{X}$, equivalently assessing $U(\boldsymbol{X})$ to be

$$U(\boldsymbol{X}) = 1 - \mathrm{E}\{m^{-1} \boldsymbol{d}(\boldsymbol{X})^T \boldsymbol{V}_X^{-1} \boldsymbol{d}(\boldsymbol{X})\}, \tag{6}$$

where $\boldsymbol{V}_X$ is the prior variance matrix of $\boldsymbol{X}$. (If $\boldsymbol{V}_X$ is of less than full rank, we first reduce $\boldsymbol{X}$ to a maximal sub-collection for which $\boldsymbol{V}_X$ has full rank.) We term (6) the *Bayes linear utility* as we may evaluate $U(\boldsymbol{X})$ purely in terms of the sufficient belief specifications necessary to carry out a Bayes linear analysis. *A priori*, our variance matrix for $\boldsymbol{d}(\boldsymbol{X})$ is $\boldsymbol{V}_X - \boldsymbol{C}_{XD} \boldsymbol{V}_D^{-1} \boldsymbol{C}_{DX}$, where $\boldsymbol{V}_D$ is the variance of the data and $\boldsymbol{C}_{XD} = \boldsymbol{C}'_{DX}$ is the covariance between $\boldsymbol{X}$ and the data, $\boldsymbol{D}$. The Bayes linear utility $U(\boldsymbol{X})$ in (6) is therefore evaluated as

$$U(\boldsymbol{X}) = 1 - m^{-1} \mathrm{trace}\{\boldsymbol{V}_X^{-1} (\boldsymbol{V}_X - \boldsymbol{C}_{XD} \boldsymbol{V}_D^{-1} \boldsymbol{C}_{DX})\} = 1 - m^{-1} \mathrm{trace}\{\boldsymbol{M}_D\} \tag{7}$$

where $\boldsymbol{M}_D = \boldsymbol{I} - \boldsymbol{V}_X^{-1} \boldsymbol{C}_{XD} V_D^{-1} \boldsymbol{C}_{DX}$ is the adjustment transform matrix for $\boldsymbol{X}$ adjusted by the data from the proposed experiment (see Farrow and Goldstein, 1993). This utility may be compared with the utility $U^{(A)}(\boldsymbol{X}) = 1 - \mathrm{E}\{m^{-1} \boldsymbol{d}(\boldsymbol{X})^T \boldsymbol{A} \boldsymbol{d}(\boldsymbol{X})\}$ which leads, in the univariate normal linear model, to the Bayesian A-optimality criterion (e.g. Chaloner and Verdinelli, 1995). In our multivariate setting we obtain $1 - m^{-1} \mathrm{trace}\{\boldsymbol{A}(\boldsymbol{V}_X - \boldsymbol{C}_{XD} \boldsymbol{V}_D^{-1} \boldsymbol{C}_{DX})\}$ and we choose $\boldsymbol{A} = \boldsymbol{V}_X^{-1}$ although other values of $\boldsymbol{A}$ could easily be used.

(iii) Suppose that we judge that increased knowledge for some linear combinations of elements of $\boldsymbol{X}$ is more important than for others. We construct the linear space $\mathcal{L}(\boldsymbol{X})$ of all linear combinations of the elements of $\boldsymbol{X}$, i.e. all quantities of the form $\sum_{i=1}^{k} c_i X_i$. Then we divide $\mathcal{L}(\boldsymbol{X})$ into linear

subspaces $\mathcal{L}_1, ..., \mathcal{L}_r$ for which we judge increased knowledge about each linear combination within a particular collection $\mathcal{L}_i$ to be of equal value. We assess the utility, $U_i(\mathcal{L}_i)$, of the experiment for learning about the elements of subspace $\mathcal{L}_i$, by applying (6) to any maximal collection of elements of $\mathcal{L}_i$ with variance matrix of full rank.

In many circumstances, we will judge gains in information over each of $\mathcal{L}_1, ..., \mathcal{L}_r$ to be mutually utility independent. Therefore, for a particular choice of design and a particular collection of experimental outcomes, we can represent our overall benefit utility, U_b, for the experiment by $(1 + k_b U_b) = \prod_{i=1}^{r} [1 + k_b b_i U_i(\mathcal{L}_i)]$, where $U_i(\mathcal{L}_i)$ is given by (6). We call this form the *mixed Bayes linear utility* for $\boldsymbol{X}$. In the special case where $k_b = 0$, U_b is given by the *additive form*

$$U_b = \sum_{i=1}^{r} b_i U_i(\mathcal{L}_i), \tag{8}$$

with constraints as in (1). Similarly, if $r = 2$, we have a binary benefit node and

$$U_b = b_1 U_1(\mathcal{L}_1) + b_2 U_2(\mathcal{L}_2) + h_b U_1(\mathcal{L}_1) U_2(\mathcal{L}_2), \tag{9}$$

with constraints as in (2), and where $r > 2$ and the parameter k_b may be nonzero,

$$U_b = \frac{\prod_{i=1}^{r} [1 + k_b b_i U_i(\mathcal{L}_i)] - 1}{\prod_{i=1}^{r} (1 + k_b b_i) - 1} \tag{10}$$

with constraints as in (3). These scalings ensure that, for any choices of trade-off parameters in the mixed Bayes linear utility, the utility of carrying out no experiment is zero, while the utility of a precise experiment for all elements of $\boldsymbol{X}$ is 1, so that the overall benefit utility is on the standard scale.

Note that, when the additive form (8) is used, this benefit utility function leads to a "weighted" or "compound" optimality criterion (e.g. Clyde and Chaloner, 1996) in which the individual terms are the Bayes linear version of the multivariate generalisation of the usual Bayesian A-optimality criterion. The forms (9) and (10) allow greater flexibility in the way that the individual "criteria" are combined.

It is also possible to justify the use of (7) as the expectation of a utility function based directly on the value of $d(X)$.

4 Choosing optimal designs for the glucose trial

4.1 Example: utilities for the glucose trial

We now illustrate the Bayes linear choice of optimal design for an experiment based on a MUIH by working through the prior assessment and design choice for the glucose trial example. For the purposes of this example, we will consider the utility of the financial costs to be proportional to the monetary costs, as given in Table 1. For simplicity, the expected marginal ethical utility for each type of patient was elicited directly as there was no prior belief in substantial correlations between benefits and ethical costs. The expected ethical costs were assessed in "units of ethical cost." (Compare Farrow and Goldstein, 1992). Both costs varied between groups. All costs were then converted to an approximately $[0, 1]$ utility scale by $U_e = (e_{max} - e_{cost})/e_{max}$ and $U_f = (f_{max} - f_{cost})/f_{max}$, where e_{cost} and f_{cost} are the ethical and financial costs and e_{max} and f_{max} are the values corresponding to a "large" experiment, actually one with 40 patients in each group. The expected ethical and financial costs (in scaled units) and the corresponding marginal utilities for a single patient in each group were as given in Table 1. Notice that the ratio of Group 2 to Group 1 costs is greater for ethical costs than for financial costs and so the design will be sensitive to the trade-off between financial and ethical costs, tending to concentrate patients in the healthier group if ethical costs are given relatively high weighting. However, the experimenters found it very hard to specify this trade-off. We will be particularly concerned with considerations of trade-off sensitivity in later sections.

Table 1: Ethical and financial costs. (1): Glucose only. (2): Glucose and C-Peptide.

	Ethical Cost	Financial Cost		$(1-U_e)\times 10^3$	$(1-U_f)\times 10^3$	
		(1)	(2)		(1)	(2)
Group 1: Well controlled	192	3.50	8.86	6.82	4.02	10.19
Group 2: Insulin	512	7.52	12.88	18.18	8.65	14.81

The benefit utility is based on our gain in knowledge. For each patient in each group, there are two possible measurements, glucose and C-peptide, at each of two time points. We wish to learn about the means at each time point for each variable in each group. Each quantity may be of clinical interest, as may be various combinations of the quantities. Because the purpose of the OGTT is to distinguish the more severe from the less severe level of diabetes, it is natural to work in terms of the following four sets of quantities which, collectively, are equivalent to the means: $\mathcal{M}_1$ and $\mathcal{M}_2$, the fasting (i.e. $t = 0$) means, and $\mathcal{M}_3$ and $\mathcal{M}_4$, the mean change over 60 minutes, for C-peptide and glucose respectively. The change in glucose level over time and the fasting level for an individual patient are diagnostic for diabetes.

These groups having been established, the marginal utility function for the benefit from learning about each of these groups of quantities is of the form (6) with $m_i = 2$. The utilities for the four groups are then combined using (8), (9) or (10) to give U_b. The financial and ethical cost utilities are combined using (5) and finally the overall utility is formed using (4).

Our overall utility function might contain products of marginal utilities and we may well have beliefs in which the utilities are not independent of each other. In some cases evaluation of expectations may require specification of moments beyond the second. However we will often be able to simplify this task, even without fully specified probability distributions. For reasons of space details are left to be presented elsewhere.

4.2 Example: choice of trade-off parameters

To simplify our account, we first consider the case where all utilities are additive. We therefore need to specify the benefit trade-off values $b_1, \ldots, b_4$ in (8), the cost trade-off a_e, a_f in (5) and the cost benefit trade-off a_c, a_b in (4).

Our experience was that the four benefit utilities were sufficiently comparable that we could ask a series of elicitation questions which effectively fixed these values. The values used for this illustration are $b_1 = 0.18$, $b_2 = 0.12$, $b_3 = 0.35$ and $b_4 = 0.35$, and these are regarded as fixed for now. (For further details of the elicitation of such a benefit utility, see Farrow and Goldstein, 1992). All evaluations of expected benefit utilities are given by (7).

The financial-ethical trade-off was assessed indirectly by comparing the desirability of omitting patients from different groups. For example, the financial cost for a patient in Group 2 is 45 % greater than for a patient in Group 1, when both variables are observed, but the ethical cost for Group 2 is 167 % greater than that for Group 1. In fact, the value chosen was $a_f = 0.45$ which suggests that one patient from Group 2 would "cost" approximately the same as two from Group 1 and gives 22% more weight to financial cost than to ethical cost in Group 1 but still gives 50% more weight to the ethical cost than to the financial cost for a patient in Group 2.

We found it helpful in eliciting the cost-benefit trade-off to evaluate the expected costs and benefits for a range of simple, standard designs, for which we chose $(n, n, 0, 0)$ designs with $n = 1, 2, \ldots$, and to elicit n_p, the preferred value of n among these, where the various costs were to be informally weighed against the square root of the expected benefit, expressed as a "reduction in standard deviation." A rough preference for $n_p = 15$, combined with $a_f = 0.45$ led to a value of $a_b = 0.85$.

The experimenters felt uncomfortable with the need to specify precise values for these trade-off parameters. Thus there is a need for methods which do not rely on such precise choice of trade-offs. We will introduce methods of this type in Section 5.

Table 2: Beliefs for simplified example.

	Mean	Variances and Covariances							
		M_{110}	M_{111}	M_{120}	M_{121}	M_{210}	M_{211}	M_{220}	M_{221}
M_{110}	0.41	0.10000	0.10000	-0.04400	-0.04400	0.05000	0.05000	-0.02200	-0.02200
M_{111}	2.17		0.20580	-0.04400	-0.08370	0.05000	0.10460	-0.02200	-0.04490
M_{120}	1.95			0.04000	0.04000	-0.02200	-0.02200	0.02000	0.02000
M_{121}	3.27				0.07145	-0.02200	-0.04490	0.02000	0.04015
M_{210}	-0.69					0.09000	0.09000	-0.03900	-0.03900
M_{211}	-0.06						0.20260	-0.03900	-0.08655
M_{220}	2.30							0.03500	0.03500
M_{221}	3.81								0.07690

	Variances and Covariances					Variances and Covariances			
	R_{1i10}	R_{1i11}	R_{1i20}	R_{1i21}		R_{2i10}	R_{2i11}	R_{2i20}	R_{2i21}
R_{1i10}	0.09905	0.06905	-0.02802	-0.02302	R_{2i10}	0.06383	0.04083	-0.00360	-0.00260
R_{1i11}		0.09905	-0.02302	-0.02802	R_{2i11}		0.06383	-0.00260	-0.00360
R_{1i20}			0.12707	0.09207	R_{2i20}			0.06383	0.04083
R_{1i21}				0.12707	R_{2i21}				0.06383

4.3 Example: prior beliefs

As our approach is based on the quantification of gains in knowledge, it is necessary to express our current state of knowledge about the O.G.T.T. Currently, elderly patients are being classified, at least in part, on their response to this test, so there already exists an implicit prior judgment as to the relevance of the outcome of the test to such patients. Establishing the framework for such a prior specification is outside the scope of this paper, and so here we give only the necessary moment specifications.

Within each group, we suppose that individuals are second-order exchangeable (i.e. second order beliefs are unaffected by permutation of the members of the group) and co-exchangeable with individuals in the other group (i.e. covariances between individuals are unaffected by permutation of members within each group). From the representation theorem for second order exchangeable structures (Goldstein, 1986), we may therefore express the observation at time t for individual i in group g for variable v as $Y_{givt} = M_{gvt} + R_{givt}$ where M_{gvt} may be considered to be the underlying *population mean* at time t in group g and R_{givt}, termed the *individual variation*, represents the deviation, from that expectation, of individual i. Individual variation of one individual is uncorrelated with that of another individual in any group and with any underlying mean. Thus $E(R_{givt}) = 0$, $E(M_{gvt}R_{g'i'v't'}) = 0$ for all $g, v, i, t, g', v', i', t'$ and $E(R_{givt}R_{g'i'v't'}) = 0$ unless $g = g'$ and $i = i'$. Thus belief specification requires prior expectations, variances and covariances for the collection of means M_{gvt} and variances and covariances within individuals for the individual variation R_{givt}.

The values are given in Table 2. Variables 1 and 2 are the logarithms of the C-peptide and glucose concentrations respectively. We worked in terms of logarithms to stabilise the variance and to make the uncertainties symmetric in the sense that the events that an observation is more than x units greater or less than its expectation would be judged to be equally likely.

4.4 Example: optimal design

Using the trade-off parameters specified in section 4.2, the optimal choice is (21, 11, 1, 0). That is, we have 22 subjects in Group 1 and 11 in Group 2 but, for 1 of the patients in Group 1, we only analyse the blood samples for glucose content. The final zero shows that it is not worth subjecting a Group 2 patient to the trial if we are not going to measure C-peptide.

We can examine the sensitivity of this choice to the trade-offs by varying the trade-off parameters. For example, holding $a_f = 0.45$ fixed but increasing a_b to 0.90, corresponding to $n_p = 20$, gives an optimum at (26,15,3,0), while decreasing a_b to 0.74, corresponding to $n_p = 10$, gives an optimum at (14,7,0,0). Similarly, holding $a_b = 0.85$ constant but varying a_f, setting $a_f = 1$, so that ethical cost is ignored, gives an optimum of (13,11,14,1), while setting $a_f = 0$, so that financial cost is ignored, gives an optimum of (24,11,0,0).

4.5 Example: generalisation to binary nodes

To illustrate the use of binary nodes we now allow h_d in (4) and h_c in (5) to be nonzero. In this example we consider U_f, U_e and U_b to be stochastically independent.

Consider first the ethical-financial trade-off. We might prefer the costs of a design where the expectations of the two scaled cost utilities, $a_f U_f$ and $a_e U_e$, are approximately equal to one where they are very different. This would suggest a positive value for h_c. With $a_f = 0.45$, $a_e = 0.55$, $h_c = 0$, we have $\mathrm{E}(a_f U_f) = \mathrm{E}(a_e U_e)$ approximately when $n_2 = 30 + 0.25 n_1$ among $(n_1, n_2, 0, 0)$ designs, for example (8,32,0,0). Also, approximately, $\mathrm{E}(a_f U_f + a_e U_e)$ is a constant when $n_1 + 2n_2$ is a constant so that, for example, we would be indifferent in terms of costs between (0,35,0,0), (8,31,0,0) and (16,27,0,0). Now consider the effect of varying h_c. If we continue to be indifferent in terms of cost between (0,35,0,0) and (16,27,0,0) this implies, approximately, that $a_e = 0.55(1 - h_c)$ and $a_f = 0.45(1 - h_c)$ but, a preference for the central design, (8,31,0,0), over either of these suggests a positive value for h_c. However, in this particular case, even increasing h_c to 1.0 is insufficient to make (8,32,0,0), a slightly larger design, preferable to (0,35,0,0) or (16,27,0,0).

If we set $h_c = 1.0$ and $a_e = a_f = 0.0$, we need to adjust a_b to 0.875 and a_c to 0.125 to preserve $n_p = 15$. The optimum design with these parameter values is (19,11,3,0). Similarly by maintaining $a_e = 0.55(1 - h_c)$ and $a_f = 0.45(1 - h_c)$ and setting $h_c = -0.81818$, the smallest value it can take under this condition, we obtain $a_e = 1.0$ and $a_f = 0.81818$. This requires adjustment of a_b to 0.81 and a_c to 0.19 and this makes the optimum design (22,11,0,0).

Restoring the financial-ethical trade-off parameters to the values specified in section 4.2, we turn our attention to h_d. If we set $h_d = 1$, we have no choice over the values of a_b and a_c. This leads to $n_p = 5$ which conflicts with the expressed value. In fact we need to reduce h_d to approximately 0.18, while holding $a_c = 0$, to restore $n_p = 15$. The overall optimum in this case is (21,11,1,0).

Similarly, if we set $h_d = -1$, we must have $a_b = a_c = 1$ and this leads to a preference for no experiment at all since this maximises the cost utility. We can restore the preference for $n_p = 15$ by setting $h_d = -0.91$, $a_b = 1.0$ and $a_c = 0.91$. This gives an overall optimum at (28,16,4,0), a design which clearly compensates for a poor cost utility with a good benefit utility.

These explorations show that, while the choice of design seems to be fairly insensitive to the value of h_c, it can be greatly affected by the value chosen for h_d.

5 Using Imprecise Trade-off Parameters

One of the most difficult tasks in specifying a mutually utility independent structure is the quantification of the various trade-off parameters in the forms (1), (2) and (3), as this typically requires the comparison of intrinsically different types of costs and benefits. For example, in medical studies, many clinicians would be unwilling to place a precise financial cost on the possible health risks to patients in a trial. Further, many experiments must satisfy various participants who may make different judgments as to such trade-offs. In practice, this means that we can often elicit U_e and U_f, but not elicit an agreed value for a_e, a_f or h_c, in (5). Therefore, it is of fundamental interest to consider problems where we are unwilling to fix on particular trade-off values or where a group of individuals must make a joint decision, and there is broad agreement on the marginal utilities, but different members of the group have different priorities when trading risks. We now develop an appropriate methodology for treating such imprecise utility.

We proceed as follows.

(i) We explain how we may formally introduce imprecision into our utility trade-offs, as follows. For each child node at which attributes are combined, we identify which preferences we are prepared

to assert between collections of sub-attribute values. From this collection of preferences, we deduce the implied constraints on the trade-off parameters which are used to combine attributes at that node. (In a precise utility specification, we would be prepared to specify preferences between all sets of attribute values, which would uniquely determine the trade-off values).

(ii) We now introduce a weak condition which is often sufficient to eliminate many of the designs under consideration, essentially that we should prefer design A to design B if we prefer A to B under every choice of trade-off parameters consistent with our stated preferences. (Under such circumstances, we would remove design B from further consideration). We then identify a greatly reduced subset of possible trade-off choices which is sufficient to check such design preferences, essentially the vertices of the region of allowable trade-off parameters identified at stage (i).

(iii) We now suggest various theoretical arguments for selecting a particular choice of design among those retained at stage (ii), which is based on optimising a linear combination of the utilities corresponding to the vertex trade-offs identified as sufficient at stage (ii). We discuss the properties of this choice, and illustrate how we may explore the sensitivity of our chosen design to variation in our specification.

If we allow imprecision in some of the elements of $\boldsymbol{\theta}$, then we refer to the resulting utility specification as an *imprecise independence hierarchy (IIH)*. This methodology is much more straightforward to describe for simple imprecise independence hierarchies, where all nodes are additive or binary, than for the more general case. If there are no multiplicative nodes in the hierarchy, then we refer to the specification as a *simple imprecise independence hierarchy (SIIH)*. To clarify the exposition, we shall first present the theory for the SIIH.

5.1 Imprecise utility trade-offs

The theory of imprecise probability can be built around the notion that, while we may be unwilling to specify a precise value for the probability of some particular event, there are various prices at which we would certainly buy a gamble on the outcome of the event and other prices at which we would certainly sell such a gamble. These preferences may be used to construct upper and lower probabilities for the corresponding event, see e.g. Walley (1991).

We may apply a similar approach to develop the theory of imprecise utility and in particular to quantify the imprecision in trade-offs between attribute values. Although we are unwilling to place strict values on the trade-offs, there will be certain combinations of outcomes over which we are prepared to state preferences and these comparisons establish the region of the space of trade-off parameters which we must consider. We choose to elicit our imprecision in the values of the trade-off parameters $\boldsymbol{\theta}$ based on our stated preferences over utility combinations for outcomes, as this is usually more meaningful than considering directly the imprecision in the elements of $\boldsymbol{\theta}$. We proceed as follows.

For each child node, we make a collection of pairwise comparisons between vectors of values of parent utilities (or, equivalently, the corresponding vectors of attribute values). As we shall vary the trade-off parameters, and thus the utilities at the child nodes, we require a scale for all utilities in the IIH, whose interpretation does not depend on the choice of trade-off parameters. Recall that the standard scale that we are using is such that a utility value of u at node i may always be interpreted as the utility of a gamble giving C_i with probability u and c_i with probability $1-u$, irrespective of the chain of trade-off parameters in the hierarchy.

At node i, we denote strict preference for utility vector $\boldsymbol{U} = (U_{1i}, U_{2i}, ..., U_{si})$ over utility vector $\boldsymbol{V} = (V_{1i}, V_{2i}, ..., V_{si})$ as $\boldsymbol{U} \succ^* \boldsymbol{V}$, and denote the weak preference, namely that we do not prefer $\boldsymbol{V}$ to $\boldsymbol{U}$ as $\boldsymbol{U} \succeq^* \boldsymbol{V}$. Each such preference places constraints on the allowable choices for the trade-off parameters $\boldsymbol{\phi}_i$. We term the collection, R, of all sets of trade-off parameters consistent with each of the stated preferences the *feasible* set of choices for the trade-off parameters. We say that the collection of pairwise comparisons is *consistent* if R is non-empty, i.e. there is at least one set of trade-off parameters satisfying all stated preferences. A comparison is *redundant* if its removal does not affect R. A change in any marginal utility will produce a change of the same sign in the overall utility. Therefore, when comparing two utility vectors where $\boldsymbol{A} \succ^* \boldsymbol{B}$, we may consider whether we can decrease any of the utilities in $\boldsymbol{A}$ or increase any of the utilities in $\boldsymbol{B}$ while

preserving the preference. In this way, we should avoid redundancy and make stringent comparisons leading to the smallest feasible set that we can determine through our stated preferences. We describe a comparison as *sharp* if $(U_{11}, \ldots, U_{s1}) \succeq^* (U_{12}, \ldots, U_{s2})$ but there is no vector $(\delta_1, \ldots, \delta_s)$ such that $\delta_i \geq 0$, for $i = 1, \ldots, s$, and $\sum_{i=1}^{s} \delta_i > 0$ for which we are prepared to assert that $(U_{11} - \delta_1, \ldots, U_{s1} - \delta_s) \succeq^* (U_{12}, \ldots, U_{s2})$. In general, we seek sharp comparisons to restrict R as far as possible.

We have described our formulation from the viewpoint of an individual who does not wish to state precise utility trade-offs. The same formulation describes a group of decision makers who agree on the marginal utilities but disagree on the utility trade-offs, where now each preference between pairs of utility vectors denotes the sharpest comparison which is agreed by all members of the group.

For additive and binary child nodes, elicitation is unconstrained. For each such node, we state whichever preferences we wish between pairs of utility vectors for the parent nodes. However, if we allow such unconstrained comparisons at the multiplicative child nodes, then the analysis that we shall develop becomes more complex. Therefore, we shall leave discussion of such nodes to section 8. We now describe the implications of the above elicitation.

Theorem 1 *The shape of the region of trade-off parameters resulting from the above elicitation scheme for an SIIH is as follows. At each additive or binary node i, we obtain a convex polyhedron R_i for the allowable values of $\boldsymbol{\phi}_i$. The regions $R_1, \ldots, R_n$ together define a region R in the combined space of parameters $\boldsymbol{\theta}$, where $\boldsymbol{\theta} \in R$ if and only if $\boldsymbol{\phi}_i \in R_i$ for $i = 1, \ldots, n$.*

Proof Consider first a single child node. We treat the two types of node in turn.

(i) Suppose node i is additive. From the definition, the parameters $a_2, \ldots, a_s$ lie within a convex polyhedron bounded by $a_2 = 0, \ldots, a_s = 0$ and $\sum_{i=2}^{s} a_i = 1$. Suppose that, for a particular pair of vectors of parent utilities $(U_{11}, \ldots, U_{s1}) \succeq^* (U_{12}, \ldots, U_{s2})$. This preference imposes the linear inequality $a_1(U_{11} - U_{12}) + a_2(U_{21} - U_{22}) + \cdots + a_s(U_{s1} - U_{s2}) \geq 0$ on the values of $a_2, \ldots, a_s$ (where $a_1 \equiv 1 - \sum_{i=2}^{s} a_i$), which is of the form $b_1 + \sum_{i=2}^{s} b_i a_i \geq 0$. If $\boldsymbol{a}_1 = (a_{12}, \ldots, a_{1s})'$ and $\boldsymbol{a}_2 = (a_{22}, \ldots, a_{2s})'$ both satisfy this inequality, then so does $\lambda \boldsymbol{a}_1 + (1 - \lambda)\boldsymbol{a}_2$. Thus the region R_i is convex and the boundary is made up of a finite number of intersecting planes.

(ii) Suppose node i is binary. If $(U_{11}, U_{21}) \succeq^* (U_{12}, U_{22})$ then $(1 - a_2 - h)(U_{11} - U_{12}) + a_2(U_{21} - U_{22}) + h(U_{11}U_{21} - U_{12}U_{22}) > 0$. Therefore, binary nodes are the same as additive nodes except that the parameters are a_2, h and the boundaries given by the definition are $a_2 = 0, a_2 = 1, h = -a_2$ and $h = 1 - a_2$.

As all preferences are based on the marginal utilities using the standard scale, the ranges of parameters at one node are independent of the values at other nodes and so the whole feasible set is as described. □

5.2 Example: trade-off imprecision

We now illustrate the construction of a feasible set. For example, in the hierarchy in the OGTT example, shown in figure 1, trade-off imprecision at node B is determined by eliciting pairwise preferences between values of the vector $\{U_1(\mathcal{L}_1), \ldots, U_4(\mathcal{L}_4)\}$, while at node C ranges are determined by preferences between vectors (E, F). In every case, preferences are expressed between lotteries over the corresponding marginal attributes. For example, at node C, we express preferences between pairs of utility vectors (E^*, F^*) and (E^{**}, F^{**}). All utilities are expressed in the standard scale.

Suppose that, at node B, for the vector $\{U_1(\mathcal{L}_1), \ldots, U_4(\mathcal{L}_4)\}$, $(0.6, 0.9, 0.7, 0.7) \succ^* (0.7, 0.7, 0.7, 0.7)$. This leads to $0.7b_1 + 0.7(1 - b_1 - b_3 - b_4) + 0.7b_3 + 0.7b_4 < 0.6b_1 + 0.9(1 - b_1 - b_3 - b_4) + 0.7b_3 + 0.7b_4$ and thence to $1.5b_1 + b_3 + b_4 < 1$. Further, increasing the value of any of the utilities in the second vector leads to uncertainty over the preference so that, for example, it is not agreed whether $(0.6, 0.9, 0.7, 0.7)$ should be preferred to $(0.71, 0.7, 0.7, 0.7)$. Therefore the comparison is sharp and we adopt $1.5b_1 + b_3 + b_4 < 1$ as part of the boundary of our parameter region. Similarly, $(0.81, 0.7, 0.7, 0.7) \succ^* (0.6, 0.74, 0.74, 0.74)$ but we are unwilling to state preferences between (0.8, 0.7, 0.7, 0.7) and (0.6, 0.74, 0.74, 0.74). The definite preference here leads to $b_1 > 0.16$.

Table 3: Trade-off parameter ranges.

	Node B				Node C			Node D		
Vertex	a_1	a_2	a_3	a_4	a_e	a_f	h_c	a_b	a_c	h_d
1	0.24	0.12	0.32	0.32	0.577	0.823	-0.400	0.82	0.16	0.02
2	0.16	0.20	0.32	0.32	0.225	0.375	0.400	0.82	0.27	-0.09
3	0.16	0.08	0.44	0.32	0.871	0.529	-0.400	0.90	0.27	-0.17
4	0.16	0.08	0.32	0.44	0.526	0.074	0.400	0.90	0.16	-0.06

Also $(0.7, 0.7, 0.77, 0.7) \succ^* (0.78, 0.78, 0.6, 0.78)$ and $(0.7, 0.7, 0.7, 0.77) \succ^* (0.78, 0.78, 0.78, 0.6)$ but we do not state a preference between (0.7,0.7,0.76,0.7) and (0.78,0.78,0.6,0.78) or between (0.7,0.7,0.7,0.76) and (0.78,0.78, 0.78,0.6). Here, the definite preferences lead to $b_3 > 0.32$ and $b_4 > 0.32$. These four inequalities define a feasible region. We could impose more constraints but the region would always be the convex hull of a finite number of vertex points. In the absence of suitable information on definite preferences, the absolute limits would apply. For example, if the first inequality above was not given, then $b_1 + b_3 + b_4 < 1$ would be the fourth face of the region.

Node C is a binary node. Suppose that, for the vector (U_e, U_f), $(0.5, 0.5) \succ^* (0.65, 0.4)$ and that this comparison is sharp, leading to $0.5a_e + 0.5a_f + 0.5 \times 0.5h_c \geq 0.65a_e + 0.4a_f + 0.65 \times 0.4h_c$ which gives $-0.15a_e + 0.1a_f - 0.01(1 - a_e - a_f) \geq 0$ and hence $a_f \leq 0.09 + 1.27a_e$. Similarly $(0.6, 0.27) \succ^* (0.5, 0.5)$ and this leads to $a_f \geq -0.62 + 1.32a_e$. To obtain two more inequalities it is helpful to introduce lotteries. Suppose that we offer a choice between the following alternatives: (1) with certainty attribute values such that $U_e = U_f = 0.5$; (2) with probability α, attribute values such that $U_e = U_f = 1$, that is zero cost, and, with probability $1 - \alpha$, attribute values such that $U_e = U_f = 0$. Suppose that (2) is preferred whenever $\alpha \geq 0.6$, leading to $0.5a_e + 0.5a_f + 0.25(1 - a_e - a_f) \leq 0.6$ and hence $a_f \leq 1.4 - a_e$, and that (1) is preferred whenever $\alpha \leq 0.4$, leading to $a_f \geq 0.6 - a_e$. These latter two constraints are equivalent to $-0.4 \leq h_c \leq 0.4$. These four constraints give the four sides of a quadrilateral region in the plane of a_e and a_f.

Consider the non-marginal node D. Suppose that we offer a choice between the following alternatives: (1) with certainty, attribute values such that $U_b = 0$ and $U_c = 1$, equivalent to the expectations when no experiment is conducted; (2) with probability α, attribute values such that $U_b = U_c = 1$, that is complete information at no cost, and, with probability $1 - \alpha$, attribute values such that $U_b = U_c = 0$, that is a costly experiment which provides no information. Suppose that (1) is preferred whenever $\alpha < 0.16$ and that (2) is preferred whenever $\alpha > 0.27$. Then $0.16 \leq a_c \leq 0.27$. Similarly, if, when offered the alternatives: (1) with certainty attribute values such that $U_b = 1$ and $U_c = 0$, that is a costly experiment which provides full information; (2) with probability α, attribute values such that $U_b = U_c = 1$ and, with probability $1 - \alpha$, attribute values such that $U_b = U_c = 0$; (1) is preferred whenever $\alpha < 0.82$ and (2) is preferred whenever $\alpha > 0.90$, then $0.82 \leq a_b \leq 0.90$. Thus we obtain a rectangular region in the plane of a_b and a_c. However, in general, the region need not be rectangular and boundaries not parallel to the a_b and a_c axes can be obtained by varying the utility values offered in the lotteries.

Table 3 gives the vertex set P_i for the feasible region R_i, at each node i.

6 Pareto optimality

6.1 Pareto optimal decisions

We have to choose from a set $\mathcal{D}$ of designs. The utility of any choice of design $A \in \mathcal{D}$ depends on the values of the trade-off parameters that we specify. As we are comparing different choices of trade-offs, we denote the utility of A, using trade-off parameters $\underline{\theta}$, evaluated as described in section 3, as $U_{A\underline{\theta}}$.

Different choices of trade-off parameters induce different preference orderings over the possible alternatives. A natural weak, partial preference ordering over allowable alternatives is that alter-

native A is at least as good as B, over feasible region R, written $A \succeq B$, if $U_{A\theta} \geq U_{B\theta}$, $\forall \boldsymbol{\theta} \in R$, A is preferred to B, over R written $A \succ B$, if $A \succeq B$ and $U_{A\theta} > U_{B\theta}$ for some $\boldsymbol{\theta} \in R$, and A is equivalent to B, written $A \simeq B$, if $U_{A\theta} = U_{B\theta} \forall \boldsymbol{\theta} \in R$. Alternative A is *Pareto optimal* for R if there is no other allowable alternative B for which $B \succ A$ over R. It seems reasonable to restrict attention to Pareto optimal alternatives. Furthermore, if we form equivalence classes of equivalent decisions $A_1 \simeq A_2 \simeq ... \simeq A_r$, then it is reasonable to restrict attention to only one representative member of each equivalence class. When we eventually choose a design, we may re-examine the collection of alternatives in the corresponding equivalence class by any subsidiary criteria, which have not yet been introduced into the problem. In this section, we discuss the identification of the Pareto optimal class. In later sections we suggest criteria for choosing between the Pareto optimal designs.

We now develop an equivalent form for the Pareto optimal designs which is much easier to compute. We first consider the difference between the expected utilities of a general pair of alternatives A, B as a function of the trade-off parameters $\boldsymbol{\theta}$, namely $d_{AB}(\boldsymbol{\theta}) = U_{A\theta} - U_{B\theta}$. We have the following lemmas.

Lemma 1 *In a SIIH utility, the utility U_j at any non-marginal node j is of the form*

$$U_j(\boldsymbol{\theta}_j) = \sum_l \left\{ \pi_{lj}(\boldsymbol{\theta}_j) U^\star_{lj} \right\} \tag{11}$$

where $\pi_{lj}(\boldsymbol{\theta}_j)$ is a product of parameters with not more than one from each node and $U^\star_{lj}$ is a product of marginal utilities (possibly just one).

Proof. Clearly (11) is true for a marginal node.

Suppose (11) is true for the utilities at each of the parent nodes of a node i. Then, whether i is additive or binary, it is clearly also true at i. Hence, by induction, (11) holds for any node. □

For each node i, we define P_i to be the set $\boldsymbol{\phi}_i = \{\phi_{i1}, \ldots, \phi_{ir(i)}\}$ of values at the $r(i) > m(i)$ vertices of R_i, and denote by P the set of overall vertex specifications for R, so that P is the subset of R with elements $\boldsymbol{\theta} = (\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_n) \in R$ such that $\boldsymbol{\phi}_i \in P_i$ for $i = 1, \ldots, n$.

Lemma 2 *If we have a SIIH utility, then, for any alternatives A, B, the minimum and maximum values of $d_{AB}(\boldsymbol{\theta})$ in R are taken at points in P.*

Proof. From Lemma 1 it follows that $d_{AB}(\boldsymbol{\theta}) = \sum_l [\pi_l(\boldsymbol{\theta}) \{ \mathrm{E}_A(U^\star_l) - \mathrm{E}_B(U^\star_l) \}]$, where E_A and E_B denote expectations under designs A, B respectively.

Consider node i, with parameters at all other nodes fixed. We see that $d_{AB}(\boldsymbol{\theta})$ is a linear function of the parameters at node i. It therefore follows that it is maximised with respect to those parameters at a vertex of R_i.

We can now fix $\phi_{i1}, \ldots, \phi_{im(i)}$ at this point in P_i and vary the parameters of another node i'. By applying the same argument we can see that we can move these parameters to a point in $P_{i'}$ without increasing $d_{AB}(\boldsymbol{\theta})$. This argument can then be applied to the parameters at each node in turn leading to the conclusion that for any point in R there is a point in P at which the value of $d_{AB}(\boldsymbol{\theta})$ is no greater.

Similarly, for any point in R there is a point in P at which the value of $d_{AB}(\boldsymbol{\theta})$ is no less. □

We now deduce that it is sufficient to check for Pareto designs in the vertex set P.

Theorem 2 *For a SIIH utility, Pareto optimal alternatives for R are the same as Pareto optimal alternatives for P.*

Proof If $d_{AB} \geq 0$ over P it follows from Lemma 2 that $d_{AB} \geq 0$ over R. The converse is obvious since $P \subseteq R$. It follows immediately that $A \succeq B$ over R if and only if $A \succeq B$ over P. If $A \succ B$ over P then $d_{AB} \geq 0$ over P and $d_{AB} > 0$ at some point in P so, since $P \subseteq R$, it also follows that $A \succ B$ over R. The converse follows since the maximum of d_{AB} in R must be in P.□

Table 4: Maxima and minima of n_{12}, n_{22}, n_{11}, n_{21} among designs optimal at the vertices of R.

	n_{12}	n_{22}	n_{11}	n_{21}
Minimum	13	8	0	0
Maximum	29	15	12	0

6.2 Example: Pareto optimal designs

For each of the sixty four vertices of R, we found the design which maximised the overall utility at this vertex. There is a wide variety among these vertex-optimal designs. Table 4 gives the maxima and minima of n_{12}, n_{22}, n_{11}, n_{21} among them.

We searched for Pareto optimal designs among all designs where $11 \leq n_{12} \leq 30$, $5 \leq n_{22} \leq 24$, $0 \leq n_{11} \leq 19$ and $0 \leq n_{21} \leq 4$. This set of 40000 candidate designs contains every design optimal at a vertex of R with some room to spare. Among these we found 433 Pareto optimal designs. The search can be conducted quickly since we know, from Theorem 2, that it is only necessary to compare candidate designs at the vertices of R.

There are various ways by which we can, in practice, further reduce the size of the Pareto optimal set, by eliminating designs which are "almost dominated" everywhere over the feasible region. For reasons of space, we shall address such reductions elsewhere, and instead, we now consider further criteria for selecting designs within the Pareto optimal collection.

7 Boundary linear utility

7.1 Definitions and motivation

The feasible region for the trade-off parameters in a SIIH is the convex hull of a finite collection of trade-off parameters $\boldsymbol{\theta}_i \in P, i = 1, \ldots, s$. Let U_i be the utility function determined by the choice of trade-offs $\boldsymbol{\theta}_i \in P, i = 1, \ldots, s$. Any function of the form

$$\bar{U}_\lambda = \sum_{i=1}^{s} \lambda_i U_i \tag{12}$$

where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_s)$ are non-negative constants such that $\sum_{i=1}^{s} \lambda_i = 1$ is termed a *boundary linear utility*. For any such $\bar{U}_\lambda$, we may identify the rule which maximises $\bar{U}_{A,\lambda} = \sum_{i=1}^{s} \lambda_i U_{A,i}$, where $U_{A,i}$ is the utility of alternative A with trade-off $\boldsymbol{\theta}_i$.

The boundary linear form is motivated by various theoretical considerations as follows. Suppose that, to solve the design problem, we intend to construct a single overall utility function $\bar{U}$. We have shown that Pareto optimality over the full feasible region of trade-off parameters is completely determined by optimality over the vertex set P. Suppose that we therefore insist that $\bar{U}$ is to be a function only of the corresponding boundary utilities $U_1, \ldots, U_s$. Each of the following arguments suggests the boundary linear form.

1. If $\bar{U}$ is to agree with the weak preference ordering, i.e. $\bar{U}(A) > \bar{U}(B)$, $\bar{U}(A) \geq \bar{U}(B)$, when $A \succ B$, $A \succeq B$, over P, respectively, then Harsanyi's theorem (Harsanyi, 1955) implies that $\bar{U}$ must be of the form (12), for some $\boldsymbol{\lambda}$. A further condition that $\bar{U}$ must be unaffected by a permutation of the utilities at the points in P would imply that $\lambda_i = 1$ for all i (see e.g. Resnik, 1987).

2. Suppose that we regard, for each alternative, the values of the boundary utilities $U_1, \ldots, U_s$ as attributes of the alternative; for example, the various boundary values of the trade-off parameters might be identified as corresponding to the different views of members of some committee which must be synthesised by the decision maker. Theorem 10.6 of Keeney and Raiffa (1976) shows that the following assumptions hold if and only if $\bar{U}$ is of the form (12):

(a) the attributes $\{U_i, U_j\}$ are preferentially independent of their complement $\tilde{U}_{ij}$ for all $i \neq j$, that is, the conditional preference structure over U_i, U_j does not depend on the other utilities;

(b) the conditional utility function $u_i^\star$ over the attribute U_i, in the combined utility, is strategically equivalent to U_i, that is, $u_i^\star$ and U_i imply the same preferences over lotteries;

(c) each attribute U_i is utility independent of the others;

(d) if there is agreement over parameter values, that is the ranges for parameters shrink to a point, then the combined utility should be the same as each of the individual utilities;

3. If we regard the boundary utilities , $U_1, \ldots, U_s$, as being the utilities of members of a group then Theorem 2 of Keeney (1976) shows that the following assumptions (adapted from Keeney) hold if and only if $\bar{U}$ is of the form (12) :

(a) $s \geq 2$, the number of alternatives is at least two and overall utilities are specified for all possible $U_1, \ldots, U_s$.

(b) If the overall utilities indicate alternative A is preferred to alternative B for a certain set of values of $U_1, \ldots, U_s$, then the overall utilities must imply that A is preferred to B if:

i. the values of $U_1, \ldots, U_s$ are not changed for alternatives other than A and

ii. the values of $U_1, \ldots, U_s$ under A either remain unchanged or are increased.

(c) If an alternative is eliminated from consideration, the new overall utilities for the remaining alternatives should be equivalent to the original overall utilities for these same alternatives.

(d) For each pair of alternatives, A and B, there is some set of values $U_1, \ldots, U_s$ such that, overall, A is preferred to B.

(e) There is no member of P such that whenever A is preferred to B at that point, A is preferred to B overall regardless of the utilities at other members of P.

In addition to such theoretical support, the boundary linear form is easy to interpret, gives a clear comparison between different choices and leads to tractable procedures even for large numbers of alternative decisions. The choice of the $\boldsymbol{\lambda}$ weights can be used to emphasise or de-emphasise the importance of a particular attribute by putting more or less weight on vertices corresponding to different values for a particular trade-off.

7.2 Properties of the boundary linear utility

From Theorem 2 we may deduce the natural relation between Pareto optimality and Bayes rules for boundary linear utilities as follows.

Corollary 1 *A decision in $\mathcal{D}$ which is either (i) a unique Bayes decision for some $\bar{U}_\lambda$, or (ii) a Bayes decision for some $\bar{U}_\lambda$, with $\lambda_i > 0$ for $i = 1, \ldots, s$, is Pareto optimal over R.*

Proof For any pair of decisions A, B, we have $\bar{U}_{A\lambda} - \bar{U}_{B\lambda} = \sum_{i=1}^{s} \lambda_i d_{AB}(\boldsymbol{\theta}_i)$. If A is the unique Bayes decision for $\bar{U}_\lambda$ then, for some i, $d_{AB}(\boldsymbol{\theta}_i) > 0$. Alternately, if A is Bayes for $\bar{U}_\lambda$, with each $\lambda_i > 0$ then, for any other decision B, either $d_{AB}(\boldsymbol{\theta}_i) = 0$ for all i, and hence, by lemma 2, $d_{AB}(\boldsymbol{\theta}) = 0$ everywhere in R, or $d_{AB}(\boldsymbol{\theta}_i) > 0$ for at least one i. Hence A is Pareto optimal over R. □

It is often helpful to equate the boundary linear form with the utility at interior trade-off values. Suppose that at node i we assign weights $\lambda_{i1}, \ldots, \lambda_{ir(i)}$ to $\boldsymbol{\phi}_{i1}, \ldots, \boldsymbol{\phi}_{ir(i)}$, for $i = 1, \ldots, n$, and that the weight λ assigned to any overall parameter specification $\boldsymbol{\theta}$ obtained by combining specifications from the n nodes, is given by the product of the corresponding weights. That is, if $\boldsymbol{\theta}$ is the combination of $\tilde{\boldsymbol{\phi}}_1, \ldots, \tilde{\boldsymbol{\phi}}_n$, which were assigned weights $\tilde{\lambda}_1, \ldots, \tilde{\lambda}_n$, at nodes $1, \ldots, n$, then $\lambda = \prod_{i=1}^{n} \tilde{\lambda}_i$. Such a weight specification is called a *multiplicative weighting*. For such a

Table 5: "Best" and "worst" Pareto optimal designs, equal vertex weights.

Rank	Design				$\bar{U}_\lambda$	Rank	Design				$\bar{U}_\lambda$
1	21	11	1	0	0.883155	11	19	11	4	0	0.883063
2	20	11	2	0	0.883146	12	22	11	1	0	0.883062
3	21	11	0	0	0.883138	426	26	15	5	0	0.878881
4	20	11	1	0	0.883137	427	23	15	9	0	0.878868
5	19	11	3	0	0.883129	428	14	8	2	0	0.878720
6	20	11	3	0	0.883115	429	13	8	4	0	0.878667
7	22	11	0	0	0.883114	430	15	8	0	0	0.878547
8	19	11	2	0	0.883114	431	13	8	3	0	0.878268
9	21	11	2	0	0.883103	432	14	8	1	0	0.878210
10	20	11	0	0	0.883073	433	13	8	2	0	0.877745

specification, we may vary the weights at each node separately. It follows directly from the fact that R_i is a convex polyhedron that, for any $\boldsymbol{\theta}$ in R, there exists a multiplicative weighting $\boldsymbol{\lambda}$ such that $\boldsymbol{\theta} = \bar{\boldsymbol{\theta}}_\lambda$ and, for any multiplicative weighting $\boldsymbol{\lambda}$, there exists a $\boldsymbol{\theta}$ in R such that $\boldsymbol{\theta} = \bar{\boldsymbol{\theta}}_\lambda$, where $\bar{\boldsymbol{\theta}}_\lambda = \sum_j \lambda_j \boldsymbol{\theta}_j$ and the sum is taken over all of the vertices of R. We now give a theorem which establishes a correspondence between the elements of R and the multiplicative boundary linear utilities.

Theorem 3 *In a SIIH, if $\boldsymbol{\lambda}$ is a multiplicative weighting then $\bar{U}_\lambda = U(\bar{\boldsymbol{\theta}}_\lambda)$.*

Proof. Consider a particular term $\pi_l(\boldsymbol{\theta})U_l^\star$ in (11). Suppose $\pi_l(\boldsymbol{\theta}) = \phi_1 \cdots \phi_M$. In $\bar{U}_\lambda$ the coefficient of $U_l^\star$ is $\sum_j \lambda_j \phi_{1k(1j)} \cdots \phi_{Mk(Mj)}$ where $k(ij)$ denotes which value of ϕ_i is selected at vertex j. Since $\boldsymbol{\lambda}$ is multiplicative, we can write $\lambda_j = \lambda_{1k(1j)} \cdots \lambda_{Mk(Mj)}$. Hence the coefficient of $U_l^\star$ can be written $\prod_m \left(\sum_k \lambda_{mk}\phi_{mk}\right)$ but, since $\sum \lambda_{nk} = 1$, we can write $\lambda_{mk} = \lambda_{mk} \prod_{n \neq m} \left(\sum_k \lambda_{nk}\right)$ and the coefficient is $\prod_m \left(\sum_j \lambda_j \phi_{mk(mj)}\right)$, which is its value in $U(\bar{\boldsymbol{\theta}}_\lambda)$. □

We know, from Theorem 3, that, for any $\boldsymbol{\theta}$ in R, we can find $\lambda_1, \ldots, \lambda_s$ such that $U(\boldsymbol{\theta}) = \sum_i \lambda_i U_i$. Values of $\boldsymbol{\theta}$ not on the boundary of R will give $\boldsymbol{\lambda}$ values satisfying the condition for Corollary 1. Rules which are Bayes for such internal $\boldsymbol{\theta}$ values will therefore be Pareto optimal over R.

7.3 Example: boundary linear design choices

Placing equal weights on all vertices gives (21,11,1,0) as the optimum design, as in Section 4.4.

Table 5 shows the first twelve and last eight designs when ranked in decreasing order of $U(\bar{\boldsymbol{\theta}}_\lambda)$ with equal weights on the vertices. Since this is a multiplicative weighting, Theorem 3 shows that $U(\bar{\boldsymbol{\theta}}_\lambda) = \bar{U}_\lambda$. There is a relatively large drop in $\bar{U}_\lambda$ after the first nine designs. The differences in utility among the first few are small. For example, an increase in U_f of 4.52×10^{-4}, that is a decrease in financial cost of 0.0452 % of f_{max}, would bring the value of $\bar{U}_\lambda$ for the tenth ranked design up to that of the ninth ranked design.

Only two of the "best nine" designs are also optimal at vertices of R, the third and seventh ranked.

Table 6 shows the maxima and minima obtained by evaluating the utility of each of the "best nine" at each vertex, subtracting it from the best utility attainable at that vertex, then multiplying by 10^3. We can make comparisons such as this because all utilities are always on the standard scale. We see that the difference for none of these designs is ever worse than 6×10^{-3} anywhere in the feasible region so there is little difference in performance between these designs. The third and seventh ranked designs, being optimal at vertices, give the best minima in Table 6. The best maximum in the table belongs to the fourth ranked design and the second best belongs to the third ranked design. The third ranked design is thus better than the first in terms of both maxima and minima. It will be noted that it is a very similar design.

Table 6: Comparison of "best nine" designs with optima at vertices. The figures shown are the maxima and minima obtained by evaluating the utility of each of the first nine designs, ranked by $U(\bar{\boldsymbol{\theta}}_\lambda)$ (equal weights), at each vertex, subtracting it from the best utility attainable at that vertex, then multiplying by 10^3.

Design rank	1	2	3	4	5	6	7	8	9
Minimum	0.041	0.134	0.000	0.095	0.066	0.140	0.000	0.079	0.116
Maximum	5.320	5.012	4.907	4.874	4.990	5.546	5.674	4.922	5.783

This analysis suggests that our design is locally robust, i.e. that there is no design of very different form to our suggested choice which is almost as good according to our suggested criterion. For reasons of space, we stop our investigation at this point. Elsewhere, we will describe how the structure allows us a much fuller investigation of the sensitivity and robustness properties of our chosen decision, based on optimisation under different choices of boundary linear utility.

8 Methodology for multiplicative nodes

We now describe how the methodology is modified for a general IIH, in which there are imprecise multiplicative nodes. In order to extend the types of analyses that we have described to a general IIH, we impose the following restrictions on the preferences that we elicit at a multiplicative child node. First, at each such node, we choose one of the parent utilities to be a standard or reference utility. Suppose that this is U_1. We set $a_1 \equiv 1$. We then assess trade-off ranges for each of the other attributes at the node by comparing that attribute with the standard. Thus, in each comparison, we vary only one of the utilities at a time, so that we compare each attribute j with attribute 1. Denote by $\boldsymbol{U}_j^+(u)$ the vector $(U_1, \ldots, U_s)$ for which $U_j = u$ and $U_i = 0, j \neq i$. For each $j > 1$, we make at least one comparison of the form $\boldsymbol{U}_1^+(u_1) \succeq^* \boldsymbol{U}_j^+(u_j)$, and optionally, several comparisons of the form $\boldsymbol{U}_j^+(u_j) \succeq^* \boldsymbol{U}_1^+(u_1)$. Secondly, we elicit pairwise preferences between utility vectors of the form $\boldsymbol{U}_1^+(u)$ and vectors $(U_{12}, \ldots, U_{s2})$, where only $U_1 = U_{12}$ and $U_j = U_{j2}$ are positive. The resulting feasible set generalises theorem 1 as follows.

Theorem 4 *The shape of the region of trade-off parameters resulting from the above elicitation scheme for an IIH is as follows. At each additive or binary node i, the shape is as given in theorem 1. For each multiplicative node i, for each fixed value of k, we obtain a bounded rectangular region $R_i(k)$ for the remaining elements of $\boldsymbol{\phi}_i$. The region R_i of allowable specifications for $\boldsymbol{\phi}_i$ is the union of the collections $\{k, R_i(k)\}$. For each fixed value of the remaining elements of $\boldsymbol{\phi}_i$, we obtain an interval for the value of k. The regions $R_1, \ldots, R_n$ together define a region R in the combined space of parameters $\boldsymbol{\theta}$, where $\boldsymbol{\theta} \in R$ if and only if $\boldsymbol{\phi}_i \in R_i$ for $i = 1, \ldots, n$.*

Proof All that we need to show is the form of the region for multiplicative nodes. Suppose node i is multiplicative. Each preference of form $\boldsymbol{U}_1^+(u_1) \succeq^* \boldsymbol{U}_j^+(u_j)$, implies $u_1 \geq a_j u_j$, so that $u_1/u_j \geq a_j$. Now suppose, for example, that the utility vector $\boldsymbol{U}_1^+(U_{11})$ is preferred to that vector where all utilities are zero except $U_1 = U_{12}$ and $U_j = U_{j2}$. This leads to $U_{11} > U_{12} + a_j U_{j2} + k a_j U_{12} U_{j2}$ and then to $k < (U_{11} - U_{12} - a_j U_{j2})/(a_j U_{12} U_{j2})$. Thus, the form for R_i is as stated. □

Note that while the feasible region for each additive and binary node is a convex polyhedron, the shape of the feasible region for a multiplicative node i is more complex. Therefore, we often choose to expand such a region to a more convenient shape $R^\star$. For additive and binary nodes we define $R_i^\star \equiv R_i$. For multiplicative nodes we define $R_i^\star$ as follows. Firstly, we identify the maximum and minimum values of k, denoted k_M, k_m, in the set R_i. For each trade-off parameter a_j, we define a_{jM}, a_{jm} to be the maximum and the minimum values of a_j respectively over the two sets $R_i(k_M), R_i(k_m)$. Denote by $R_i^\star$, the rectangular region bounded by the values $k_m < k < k_M, a_{jm} < a_j < a_{jM}, j = 2, ...$ Then $R_i^\star$ is the smallest rectangular region for which $R_i \subseteq R_i^\star$. When we replace each such R_i by $R_i^\star$, we denote the corresponding extended feasible region as

$R^\star$. For each additive or binary node define $P_i^\star$ to be the set $\{\boldsymbol{\phi}_{i1}, \ldots, \boldsymbol{\phi}_{ir(i)}\}$ of the $r(i) > m(i)$ vertices of $R_i^\star$. For each multiplicative node with s parents define $P_i^\star$ to be the set of $\boldsymbol{\phi}_i$ such that $a_j = a_{jm}$ or $a_j = a_{jM}$, for $j = 2, \ldots, s$, and $k_m < k < k_M$. Then $P^\star$ is the set of overall specifications each element of which is the combination of one element from each of $P_1^\star, , \ldots, P_n^\star$. We may now generalise theorem 2 as follows. The details of the proof are in the appendix.

Theorem 5 *For a general IIH utility, Pareto optimal alternatives for $R^\star$ are the same as Pareto optimal alternatives for $P^\star$.*

The further modification we must make is to the boundary linear forms. If there are some multiplicative nodes with imprecise values for the parameter k in (2), the corresponding boundary is continuous over k. The natural generalisation of the boundary linear form to a continuous weight function is as follows. Let the imprecise k parameters be $\boldsymbol{k} = (k_1, \ldots, k_w)'$. Now let $\boldsymbol{\theta}_i$ denote a vertex specification of the other parameters so that each element of P is a particular $\boldsymbol{\theta}_i$ combined with a value for $\boldsymbol{k}$ and replace λ_i with $\lambda_i(\boldsymbol{k})$, a scaled probability density function for $\boldsymbol{k}$ so that

$$\sum_{i=1}^{s} \int \cdots \int \lambda_i(\boldsymbol{k}).dk_1 \ldots dk_w = 1,$$

where the integrals are taken over the ranges of $k_1, \ldots, k_w$. Then

$$\bar{U}_\lambda = \sum_{i=1}^{s} \int \cdots \int \lambda_i(\boldsymbol{k}) U_i.dk_1 \ldots dk_w.$$

We may also allow $\lambda_i(\boldsymbol{k})$ to be wholly or partly discrete with some or all of the weight concentrated at discrete values of some or all of $k_1, \ldots, k_w$. In this case, integrals are replaced by summations as appropriate. We may generalise theorem 3 as follows. The proof is given in the appendix.

Theorem 6 *In any IIH, for any $\boldsymbol{\theta}$ in R there exists a multiplicative weighting $\boldsymbol{\lambda}$ such that $U(\boldsymbol{\theta}) = \bar{U}_\lambda$.*

9 Concluding comments

This paper has two aims. Firstly, we show how mutually utility independent hierarchies, which weigh the various costs of the experiment against benefits expressed through a mixed Bayes linear utility, provide a flexible and intuitive methodology for experimental design which remains tractable even for complex multi-variate problems. Secondly, we develop methodology for problems in which the experimenters are unwilling to specify unique trade-off values between the attributes in the hierarchy. Partly, we do this as we consider that precise specification of such trade-offs is likely to be one of the main stumbling points in implementing our approach. Partly also, we consider that problems in imprecise utility are of fundamental interest in their own right, in much the same way as are problems in imprecise probability.

This paper is intended to offer a framework for experimental design which respects the complexities of the cost and benefit structure. The imprecise utility formulation is intended to identify classes of designs which are worthy of further consideration and to suggest criteria for selecting between these designs. This raises important methodological questions concerning the best ways to identify and choose between the leading contenders for the best choice of design, and these questions will be pursued elsewhere.

APPENDIX: PROOFS OF THEOREMS

Proof of Theorem 5

First we need to modify Lemma 1. We now show that

$$U(\boldsymbol{\theta}) = \sum_l \{\pi_l^\star(\boldsymbol{\theta}) U_l^\star\}, \tag{13}$$

where $\pi_l^\star(\boldsymbol{\theta}) = \pi_{l1}^\star(\boldsymbol{\theta})/\pi_{l2}^\star(\boldsymbol{\theta})$.

Here $\pi_{l1}^\star(\boldsymbol{\theta})$ is a product of parameters with not more than one from each additive or binary node and any parameters, other than k, from a multiplicative node which are present are present linearly and $\pi_{l2}^\star(\boldsymbol{\theta})$ is either a product of terms of the form $\prod_{i=1}^s (1 + ka_i) - 1$ from multiplicative nodes or is 1.

Clearly (13) holds for a marginal node.

Suppose (13) holds for the utilities at each of the parent nodes of a node i. Then, if i is additive or binary, it clearly also holds at i. If i is multiplicative then we observe that the numerator in (3) can be written as a sum of terms, each of which is of the form $k^m \prod a_j^{p(j)} U_j^{p(j)}$ where $p(j)$ is either 0 or 1 and U_j is a parent utility. It follows that (13) holds for node i. Hence, by induction, (13) holds for the overall utility.

Now we can write $d_{AB}(\boldsymbol{\theta}) = \sum_l [\pi_l^\star(\boldsymbol{\theta}) \{\mathrm{E}_A(U_l^\star) - \mathrm{E}_B(U_l^\star)\}]$.

Consider a single parameter a_{ij} at a multiplicative node i. We can write $d_{AB}(\boldsymbol{\theta})$ in the form $f(a_{ij}) = [C_1 + C_2 a_{ij}]/[C_3 + C_4 a_{ij}] + C_5$, where $C_1, \ldots, C_5$ are constants. As $\partial f(a_{ij})/\partial a_{ij} = (C_2 C_3 - C_1 C_4)/(C_3 - C_4 a_{ij})^2$, it follows that $f(a_{ij})$ is monotonic in a_{ij}, over any domain in which $|f(a_{ij})|$ is bounded away from infinity. So $d_{AB}(\boldsymbol{\theta})$ is monotonic in any single a_{ij}. If we minimise $d_{AB}(\boldsymbol{\theta})$ with respect to any single a_{ij} we will always reach one of the limits for a_{ij}. Hence the minimum of $d_{AB}(\boldsymbol{\theta})$ for any fixed k_i must be at a vertex of the subset of $R_i^\star$, for this value of k_i, where $R_i^\star$ for a multiplicative node is the rectangular region bounded by the upper and lower limits of $a_{i2}, \ldots, a_{is(i)}$. Now, if we allow k_i to vary, we see that the global minimum of $d_{AB}(\boldsymbol{\theta})$ must be at a point in $P_i^\star$.

The rest of the proof follows as before.

Proof of Theorem 6

Suppose that multiplicative weightings can be found for each of the parent nodes of node i such that $U(\boldsymbol{\theta}_j) = \bar{U}_{\lambda j}$, where the subscript j denotes that this applies to the sub-hierarchy under node j. Trivially we can find such a weighting for a marginal node. If node i is additive or binary it is clear from Theorem 3 that we can find a multiplicative weighting such that $U(\boldsymbol{\theta}_i) = \bar{U}_{\lambda i}$.

Suppose node i is multiplicative. Our definition allows us to concentrate all of the weight at the chosen value of k_i so we can now consider k_i to be fixed. Suppose that we assign weights $\lambda_{ij}^\star$ and $(1 - \lambda_{ij}^\star)$ to the lower and upper values of $a_{i2}, \ldots, a_{ir(i)}$, where $a_{i1} = 1 - \sum_{j=2}^{r(i)} a_{ij}$, and then obtain weights for the vertices of R_i by multiplying together the appropriate combinations of these. In this way we can consider averaging over each of $a_{i2}, \ldots, a_{ir(i)}$ in turn. The averaging property will now apply, with respect to $a_{i2}, \ldots, a_{ir(i)}$, to the unscaled utility

$$U_i^\star = k_i^{-1} \left[\prod_{j=1}^{r(i)} (1 + k_i a_{ij} U_{ij}) \right] - 1 = k_i^{-1} (1 + k_i a_{ij} U_{ij}) \left[\prod_{f \neq j} (1 + k_i a_{if} U_{if}) \right] - 1$$

where $U_i = U_i^\star / F_i$ and

$$F_i = k_i^{-1} \left[\prod_{j=1}^{r(i)} (1 + k_i a_{ij}) \right] - 1 = k_i^{-1} (1 + k_i a_{ij}) \left[\prod_{f \neq j} (1 + k_i a_{if} U_{if}) \right] - 1$$

but does not apply exactly to the scaled utility U_i because F_i^{-1} is not linear in a_{ij}.

Now write $U_i(a_{ij})$ and $F_i(a_{ij})$ to show dependence on a_{ij} and let the lower and upper values of a_{ij} be a_{ij1}, a_{ij2} (such that $a_{ij1} < a_{ij2}$). Choose any other value a_{ij3} with $a_{ij1} \leq a_{ij3} \leq a_{ij2}$. We can write $a_{ij3} = \mu_{ij} a_{ij1} + (1 - \mu_{ij}) a_{ij2}$ for some μ_{ij} with $0 \leq \mu_{ij} \leq 1$. Clearly

$$\begin{aligned} U_i(a_{ij3}) &= [U_i^\star(\mu_{ij}a_{ij1} + \{1-\mu_{ij}\}a_{ij2})]/[F_i(\mu_{ij}a_{ij1} + \{1-\mu_{ij}\}a_{ij2})] \\ &= [\mu_{ij}U_i^\star(a_{ij1}) + \{1-\mu_{ij}\}U_i^\star(a_{ij2})]/[\mu_{ij}F_i(a_{ij1}) + \{1-\mu_{ij}\}F_i(a_{ij2})] \end{aligned}$$

but

$$\lambda_{ij}^\star U_i(a_{ij1}) + (1-\lambda_{ij}^\star)U_i(a_{ij2}) = \lambda_{ij}^\star[U_i^\star(a_{ij1})]/[F_i(a_{ij1})] + (1-\lambda_{ij}^\star)[U_i^\star(a_{ij2})]/[F_i(a_{ij2})].$$

We require

$$[\mu_{ij}]/[\mu_{ij}F_i(a_{ij1}) + (1-\mu_{ij})F_i(a_{ij2})] = [\lambda_{ij}^\star]/[F_i(a_{ij1})]$$

and

$$[(1-\mu_{ij})]/[\mu_{ij}F_i(a_{ij1}) + (1-\mu_{ij})F_i(a_{ij2})] = [(1-\lambda_{ij}^\star)]/[F_i(a_{ij2})]$$

and these are satisfied when

$$\lambda_{ij}^\star = [\mu_{ij}F_i(a_{ij1})]/[\mu_{ij}F_i(a_{ij1}) + (1-\mu_{ij})F_i(a_{ij2})]$$

which is in $0 \leq \lambda_{ij}^\star \leq 1$.

We can now extend the argument to the other parameters at node i.

Thus we can find a multiplicative weighting such that $U(\boldsymbol{\theta}_i) = \bar{U}_{\lambda i}$ and, by induction, we can find a multiplicative weighting such that $U(\boldsymbol{\theta}) = \bar{U}_\lambda$.

□

References

Adcock, C.J., 1997. Sample size determination: a review. The Statisticican, 46, 261-283.

Bernardo, J.M., 1997. Statistical inference as a decision problem: the choice of sample size. The Statisticican, 46, 151-153.

Chaloner, K., and Verdinelli, I., 1995. Bayesian experimental design, a review. Statistical Science 10, 273-304.

Claxton, K., Lacey, L.F., and Walker, S.G., 2000. Selecting treatments: a decision theoretic approach. Journal of the Royal Statistical Society, Ser. A, 163, 211-225.

Clyde, M. and Chaloner, K., 1996. The equivalence of constrained and weighted designs in multiple objective design problems. Journal of the American Statistical Association, 91, 1236-1244.

Dette, H. and Franke, T., 2000. Constrained $D-$ and D_1-optimal designs for polynomial regression. The Annals of Statistics, 28, 1702-1727.

Farrow, M. and Goldstein, M., 1992. Reconciling costs and benefits in experimental design. In: Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M. (Eds.), Bayesian Statistics 4, Oxford University Press, Oxford, 607-615.

Farrow, M. and Goldstein, M., 1993. Bayes linear methods for grouped multivariate repeated measures studies with application to crossover trials. Biometrika, 80, 39-59.

Farrow, M. and Leyland, A.H., 1991. Interpretation of oral glucose tolerance test results. In: Dunstan,F., Pickles,J. (Eds.), Statistics in Medicine, Clarendon Press, Oxford, 249-266.

de Finetti, B., 1974. Theory of Probability: A Critical Introductory Treatment, Volume 1. John Wiley, New York.

Gittins, J. and Pezeshk, H., 2000. How large should a clinical trial be? The Statistician, 49, 177-187.

Goldstein, M., 1986. Exchangeable belief structures. Journal of the American Statistical Association, 81, 971-976.

Goldstein, M., 1999. Bayes linear analysis. In: Kotz, S., Read,C., Banks, D.L. (Eds.), Encyclopedia of Statistical Sciences, Update volume 3. John Wiley, Chichester, 29-34.

Goldstein, M., Farrow, M. and Spiropoulos, T., 1993. Prediction under the influence: Bayes linear influence diagrams for prediction in a large brewery. The Statistician, 42, 445-459.

Goldstein, M. and Wooff, D.A., 1997. Choosing sample sizes in balanced experimental designs: a Bayes linear approach. The Statistician, 46, 167-183.

Harsanyi, J.C., 1955. Cardinal welfare, individualistic ethics and interpersonal comparisons of utility. Journal of Political Economy, 63, 309-321.

Keeney, R.L., 1976. A group preference axiomatization with cardinal utility. Management Science, 23, 140-145.

Keeney, R.L. and Raiffa, H., 1976. Decisions with Multiple Objectives: Preferences and Value Trade-offs.John Wiley & Sons, New York.

Lindley, D.V., 1997. The choice of sample size. The Statistician, 46, 129-138.

Müller, P., 1999. Simulation-based optimal design. In: Bernardo,J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M. (Eds.), Bayesian Statistics 6, Oxford University Press, Oxford, 459-474.

Resnik, M.D., 1987. Choices: an Introduction to Decision Theory. University of Minnesota Press, Minneapolis.

Shaw, S.C. and Goldstein, M., 1999. Simplifying complex designs: Bayes linear experimental design for grouped multivariate exchangeable systems. In: Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M. (Eds.), Bayesian Statistics 6, Oxford University Press, Oxford, 839-848.

Tan, S.B. and Smith, A.F.M., 1998. Exploratory thoughts on clinical trials with utilities. Statistics in Medicine, 17, 2771-2791.

Verdinelli, I. and Kadane, J.B., 1992. Bayesian designs for maximising information and outcome. Journal of the American Statistical Association, 87, 510-515.

Walker, S.G., 2003. How many samples?: a Bayesian nonparametric approach. The Statistician, 52, 475-482.

Walley, P., 1991. Statistical Reasoning with Imprecise Probabilities. Chapman and Hall, London.

World Health Organisation, 1980. W.H.O. Expert Committee on diabetes mellitus. World Health Organisation Technical Report Series 646, W.H.O., Geneva.

Wickramasinghe, L.S.P., Chazan, B.I., Farrow, M., Bansal, S.K. and Basu, S.K., 1992. C-peptide response to oral glucose and its clinical role in elderly people. Age and Ageing, 21, 103-108.